

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Прикладной искусственный интеллект	Код ОП 09.03.03
Направление подготовки Прикладная информатика	Код направления и уровня подготовки 09.03.03

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Аксёнов Александр Сергеевич	без ученой степени, без ученого звания	Ассистент	информационных технологий и систем управления
2	Ронкин Михаил Владимирович	к.т.н.	Доцент	информационных технологий и систем управления
3	Созыкин Андрей Владимирович	к.т.н.	Доцент	информационных технологий и систем управления

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ **Пакетная и потоковая обработка данных**

1.1. Аннотация содержания модуля

Потоковая и пакетная обработка данных» направлен на формирование у студентов компетенций в области подготовки данных для дальнейшей аналитики как в режиме онлайн (OLAP-системы), так и в системах близких к реальному времени (Near Real Time). В системах хранилищ и озер данных процессы захвата изменений могут происходить как пакетно (batch processing), так и потоково (streaming processing). Студент в ходе обучения научится понимать отличия этих процессов, особенности и архитектуру. Рассматриваются OLAP и брокер системы на примере Spark, kafka, Flink.

1.2. Структура и объем модуля

Таблица 1

№	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах
1	Пакетная и потоковая обработка данных	4
ИТОГО по модулю:		4

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	Не предусмотрены
Постреквизиты и кореквизиты модуля	Не предусмотрены

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
1	2	3	4
Пакетная и потоковая обработка данных	ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-6.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения	ПК-6.2. 3-1. Знает методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных ПК-6.2. 3-2. Знает методы планирования вычислительного эксперимента, формирования

			<p>обучающей и контрольной выборок</p> <p>ПК-6.2. У-1. Умеет выявлять и исключать из массива данных ошибочные данные и выбросы</p> <p>ПК-6.2. У-2. Умеет выделять входные и выходные переменные с целью использования предиктивных моделей</p> <p>ПК-6.2. У-3. Умеет осуществлять разметку структурированных и неструктурированных данных</p> <p>ПК-6.2. У-4. Умеет использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения</p> <p>ПК-6.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных</p>
<p>Пакетная и потоковая обработка данных</p>	<p>ПК-7. Способен разрабатывать системы анализа больших данных</p>	<p>ПК-7.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных</p>	<p>ПК-7.1 3-1. Знает общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных</p> <p>ПК-7.1 3-2. Знает принципы работы экосистемы Hadoop, фреймворка SPARK</p> <p>ПК-7.1 3-3. Знает устройство интерфейсов между реляционными SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных</p> <p>ПК-7.1 3-4. Знает предметно-ориентированные языки</p> <p>ПК-7.1 У-1. Умеет настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных</p> <p>ПК-7.1 У-2. Умеет разрабатывать программное обеспечение для очистки и валидации наборов больших данных</p> <p>ПК-7.1 У-3. Умеет выполнять потоковую обработку данных (data streaming, event processing)</p> <p>ПК-7.1 У-4. Умеет использовать шины данных (Apache Kafka)</p>

		ПК-7.1 У-5. Умеет использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа) и операций с большими данными (например, матричные операции)
	ПК-7.2. Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	<p>ПК-7.2. З-1. Знает принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных</p> <p>ПК-7.2. З-2. Знает устройство и принципы работы систем обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL)</p> <p>ПК-7.2. З-3. Знает архитектуру и принципы работы промышленных решений, созданных на основе искусственного интеллекта</p> <p>ПК-7.2. З-4. Знает методы и технологии машинного обучения на больших данных</p> <p>ПК-7.2. У-1. Умеет разрабатывать программное обеспечение для анализа больших данных</p> <p>ПК-7.2. У-2. Умеет разрабатывать программные и технические средства визуализации больших данных и результатов их анализа</p> <p>ПК-7.2. У-3. Умеет использовать системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL, процессы и инструменты)</p> <p>ПК-7.2. У-4. Умеет использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности</p> <p>ПК-7.2. У-5. Умеет описывать и управлять качеством и достоверностью больших данных</p>

1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной форме.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Пакетная и потоковая обработка данных

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Аксёнов Александр Сергеевич	без ученой степени, без ученого звания	Ассистент	информационных технологий и систем управления
2	Ронкин Михаил Владимирович	к.т.н.	Доцент	информационных технологий и систем управления
3	Созыкин Андрей Владимирович	к.т.н.	Доцент	информационных технологий и систем управления

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

Авторы:

- Аксёнов Александр Сергеевич, Ассистент, информационных технологий и систем управления
- Ронкин Михаил Владимирович, Доцент, информационных технологий и систем управления
- Созыкин Андрей Владимирович, Доцент, информационных технологий и систем управления

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - Базовый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания; Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
1	Распределенные данные, распределенная обработка и современные требования к скорости и надежности вычислений	Понятие Big Data и его составляющие. Требования к высоконагруженным сервисам. Модели данных и языки запросов. Распределенные хранилища и распределенные данные. Производные данные. Введение в понятие пакетной и потоковой обработки
2	Spark. Основные парадигмы. Стадии обработки. Типы операций. API взаимодействия	Что такое spark? Место в экосистеме больших данных. Модель параллельных вычислений: отложенные вычисления, хранение данных в памяти и управление памятью, неизменяемость и интерфейс RDD, широкие и узкие зависимости, планирование заданий.
3	Spark. Работа с dataframe	Сессии и контексты. Схемы. API DataFrame. Оптимизация запросов. Эффективные преобразования.
4	Работа с данными типа "ключ — значение" (rdd)	Как использовать классы PairRDDFunctions и OrderedRDDFunctions. Действия над парами "ключ — значение". Выбор операции агрегирования. Предотвращение

		ошибок нехватки памяти при операциях агрегирования. Методы секционирования и данные типа "ключ — значение". Выявление отстающих задач и несбалансированных данных
5	Spark. Обработка данных и ML	Выбор между библиотеками Spark MLlib и Spark ML. Кодирование признаков в библиотеке MLlib и подготовка данных. Обучение моделей библиотеки MLlib. Оценка модели MLlib. Этапы конвейера ML и MLlib
6	Очереди и брокеры	Суть понятий. Обзор брокеров. Kafka. Основные понятия, архитектура, пакеты исходного кода. Разработка проекта, производители и потребители. Брокеры и темы. Kafka, как хранилище. Kafka Streams
7	Потоковая обработка с учетом состояний и основы потоковой обработки	Flink: архитектура, DataStream, операторы на основе времени и оконные операторы. Обработка потоков с учетом состояния. Архитектура Apache Flink Настройка Flink для потоковых приложений

1.3. Направление, виды воспитательной деятельности и используемые технологии

Таблица 1.2

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения
Профессиональное воспитание	профориентационная деятельность	Технология самостоятельной работы Технология анализа образовательных задач	ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-6.2. Умеет использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения

1.4. Программа дисциплины реализуется на государственном языке Российской Федерации.

2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Пакетная и потоковая обработка данных

Электронные ресурсы (издания)

1. Клеппман М., Высоконагруженные приложения. Программирование, масштабирование, поддержка. — СПб.: Питер, 2018. — 640 с.

2. Стренгхолт Питхейн, Масштабируемые данные. Лучшие шаблоны высоконагруженных архитектур. — СПб.: Питер, 2022. — 368 с
3. Бёрнс Б. Распределенные системы. Паттерны проектирования. — СПб.: Питер, 2019. — 224 с
4. Дэвис К. Шаблоны проектирования для облачной среды / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 388
5. Уэске Ф., Калаври В. Поточковая обработка данных с Apache Flink / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2021. – 298
6. Дилан Скотт, Виктор Гамов, Дейв Клейн, Kafka в действии / пер. с англ. А. Н. Киселева. – М.: ДМК Пресс, 2022. – 310 с
7. Нархид Ния, Шапира Гвен, Палино Тодд, Apache Kafka. Поточковая обработка и анализ данных. — СПб.: Питер, 2019. — 320 с
8. Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс, Spark для профессионалов: современные паттерны обработки больших данных. — СПб.: Питер, 2017. — 272 с
9. Холден Карау, Рейчел Уоррен, Эффективный Spark. Масштабирование и оптимизация. — СПб.: Питер, 2018. — 352 с

Профессиональные базы данных, информационно-справочные системы

1. Единое окно доступа к образовательным ресурсам. Раздел Информатика и информационные технологии <http://window.edu.ru/catalog>
2. Интернет-Университет Информационных Технологий <http://www.intuit.ru/>
3. Веб-сервис для хостинга IT-проектов и их совместной разработки Github <http://www.github.ru>

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

1. ЭБС Университетская библиотека онлайн «Директ-Медиа» <http://www.biblioclub.ru/>
2. eLibrary ООО Научная электронная библиотека <http://elibrary.ru>

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Пакетная и потоковая обработка данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения
1	Лекции	Мебель аудиторная с количеством рабочих мест в	Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc

		<p>соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p> <p>Подключение к сети Интернет</p>	<p>Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p>
2	Лабораторные занятия	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p> <p>Периферийное устройство</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc</p> <p>Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p>
3	Самостоятельная работа студентов	<p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc</p> <p>Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p>
4	Текущий контроль и промежуточная аттестация	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc</p> <p>Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p>

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Пакетная и потоковая обработка данных

Код модуля
1164092(1)

Модуль
Пакетная и потоковая обработка данных

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Аксёнов Александр Сергеевич	-	ассистент	информационных технологий и систем управления
2	Ронкин Михаил Владимирович	к.т.н.	Доцент	информационных технологий и систем управления
3	Созыкин Андрей Владимирович	к.т.н.	Доцент	информационных технологий и систем управления

Авторы:

- Аксёнов Александр Сергеевич, ассистент, Информационных технологий и систем управления
- Ронкин Михаил Владимирович, Доцент, информационных технологий и систем управления
- Созыкин Андрей Владимирович, Доцент, информационных технологий и систем управления

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ *Пакетная и потоковая обработка данных*

1.	Объем дисциплины в зачетных единицах	4	
2.	Виды аудиторных занятий	Лекции Лабораторные занятия	
3.	Промежуточная аттестация	Экзамен	
4.	Текущая аттестация	Контрольная работа	1
		Домашняя работа	1

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ *МОДУЛЯ* *Пакетная и потоковая обработка данных*

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3	4
ПК-6. Способен осуществлять сбор и подготовку данных для систем	ПК-6.2. Выполняет подготовку и разметку структурированы	ПК-6.2. 3-1. Знает методы редукиции размерности элементов набора данных и их предварительной статистической обработки,	Домашняя работа Экзамен Контрольная работа

искусственного интеллекта	х и неструктурированных данных для машинного обучения	разметки структурированных и неструктурированных данных ПК-6.2. З-2. Знает методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок ПК-6.2. У-1. Умеет выявлять и исключать из массива данных ошибочные данные и выбросы ПК-6.2. У-2. Умеет выделять входные и выходные переменные с целью использования предиктивных моделей ПК-6.2. У-3. Умеет осуществлять разметку структурированных и неструктурированных данных ПК-6.2. У-4. Умеет использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения ПК-6.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных	Лабораторные занятия Лекции
ПК-7. Способен разрабатывать системы анализа больших данных	ПК-7.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной	ПК-7.1 З-1. Знает общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных ПК-7.1 З-2. Знает принципы работы экосистемы Hadoop, фреймворка SPARK ПК-7.1 З-3. Знает устройство интерфейсов между реляционными SQL-хранилищами данных и	Домашняя работа Экзамен Контрольная работа Лабораторные занятия Лекции

	<p>архитектуры больших данных</p>	<p>нереляционными NoSQL-хранилищами данных ПК-7.1 З-4. Знает предметно-ориентированные языки ПК-7.1 У-1. Умеет настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных ПК-7.1 У-2. Умеет разрабатывать программное обеспечение для очистки и валидации наборов больших данных ПК-7.1 У-3. Умеет выполнять потоковую обработку данных (data streaming, event processing) ПК-7.1 У-4. Умеет использовать шины данных (Apache Kafka) ПК-7.1 У-5. Умеет использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа) и операций с большими данными (например, матричные операции)</p>	
	<p>ПК-7.2. Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных</p>	<p>ПК-7.2. З-1. Знает принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных ПК-7.2. З-2. Знает устройство и принципы работы систем обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL) ПК-7.2. З-3. Знает архитектуру и принципы работы промышленных решений, созданных на</p>	<p>Домашняя работа Экзамен Контрольная работа Лабораторные занятия Лекции</p>

		<p>основе искусственного интеллекта</p> <p>ПК-7.2. 3-4. Знает методы и технологии машинного обучения на больших данных</p> <p>ПК-7.2. У-1. Умеет разрабатывать программное обеспечение для анализа больших данных</p> <p>ПК-7.2. У-2. Умеет разрабатывать программные и технические средства визуализации больших данных и результатов их анализа</p> <p>ПК-7.2. У-3. Умеет использовать системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL, процессы и инструменты)</p> <p>ПК-7.2. У-4. Умеет использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности</p> <p>ПК-7.2. У-5. Умеет описывать и управлять качеством и достоверностью больших данных</p>	
--	--	---	--

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.6		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	6,7	50
<i>контрольная работа</i>	6,14	50
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – экзамен		

Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям–нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –0.4		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>отчет по лабораторным работам</i>	6,16	100
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям -1		
Промежуточная аттестация по лабораторным занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)

2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Лабораторные занятия

Примерный перечень тем

1. Установка и развертывание Apache Spark
2. Применение Apache Spark для считывания, обработки и записи данных
3. Применение data frame API Apache Spark
4. Применение rdd API Apache Spark
5. Применение Apache Spark для решения задачи преобразования данных
6. Применение Apache Spark для решения задачи анализа данных
7. Применение Apache Spark для решения задачи машинного обучения
8. Установка и развертывание Apache Kafka
9. Создания простейшего сервиса Apache Kafka, который «слушает» источник и передает данные на Apache Spark job
10. Установка и развертывание Apache Flink
11. Создание простейшего сервиса Apache Flink поставляющего данные на основе состояний

LMS-платформа – не предусмотрена

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа

Примерный перечень тем

1. Распределенное хранение и обработка данных

Примерные задания

1. Опишите модель параллельных вычислений фреймворка Spark:

2. Опишите преимущества и недостатки отложенных вычислений. Отказоустойчивость

3. Как происходит хранение данных в памяти и управление памятью

4. Опишите понятия неизменяемости и интерфейс RDD

5. Какие типы наборов RDD вы знаете

6. Чем отличаются преобразования и действия?

7. Чем отличаются широкие и узкие зависимости?

LMS-платформа – не предусмотрена

5.2.2. Домашняя работа

Примерный перечень тем

1. Особенности применения методов распределенных вычислений

Примерные задания

Цель работы:

Изучить возможности Apache Spark и выделить ключевые особенности использования данного инструмента для чтения, преобразования, записи данных, а также решения задач анализа данных

Задание и требования:

Домашняя работа включает изучение литературы и документации по Apache Spark, а также применение изученных методов и функций для решения практической задачи на выданном наборе данных. Задание будет включать следующие этапы: считать набор данных, выполнить необходимые преобразования и расчеты, сохранить преобразованный набор данных.

Работа должна быть оформлена в виде jupyter notebook или python скрипта и отправлена в форму для приема работы. Задание индивидуально

LMS-платформа – не предусмотрена

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Экзамен

Список примерных вопросов

1. Что такое Spark?

2. Каковы ключевые особенности Spark?

3. Что такое SCC?

4. Что такое RDD?

5. Что такое неизменяемость (Immutability)?

6. Что такое YARN?

7. Какие самые распространённые языки программирования в Spark?

8. Сколько менеджеров кластера доступны в Spark?
 9. Каковы обязанности движка Spark?
 10. Что такое ленивые вычисления?
 11. Что такое раздел (Partition)?
 12. Для чего нужен Spark Streaming?
 13. Нормально ли запускать все ваши процессы на локализованном ноде?
 14. Для чего используется SparkCore?
 15. Имеет ли применение File System API в Spark?
 16. Чем MapReduce отличается от Spark?
 17. Что вы понимаете под трансформациями в Spark?
 18. Что такое Apache Kafka?
 19. Как запустить сервер в Kafka?
 20. Что такое традиционные методы передачи данных и чем Kafka лучше?
 21. Что такое zookeeper в Kafka и можем ли мы использовать эту программу без него?
 22. Почему Kafka является такой важной частью технологии?
 23. Объясните, что такое последователь и лидер в Kafka.
 24. Что такое потребители и пользователи в Kafka?
 25. Как вы используете Kafka в качестве системы хранения данных?
 26. Объясните максимальный размер сообщения, которое может принять Kafka.
 27. Как разбалансировать кластер в Kafka?
- LMS-платформа – не предусмотрена

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения	Контрольно-оценочные мероприятия
Профессиональное воспитание	профорориентац ионная деятельность	Технология самостоятельн ой работы Технология анализа образовательн ых задач	ПК-6	ПК-6.2. У- 4	Домашняя работа Экзамен Контрольная работа Лабораторны е занятия Лекции